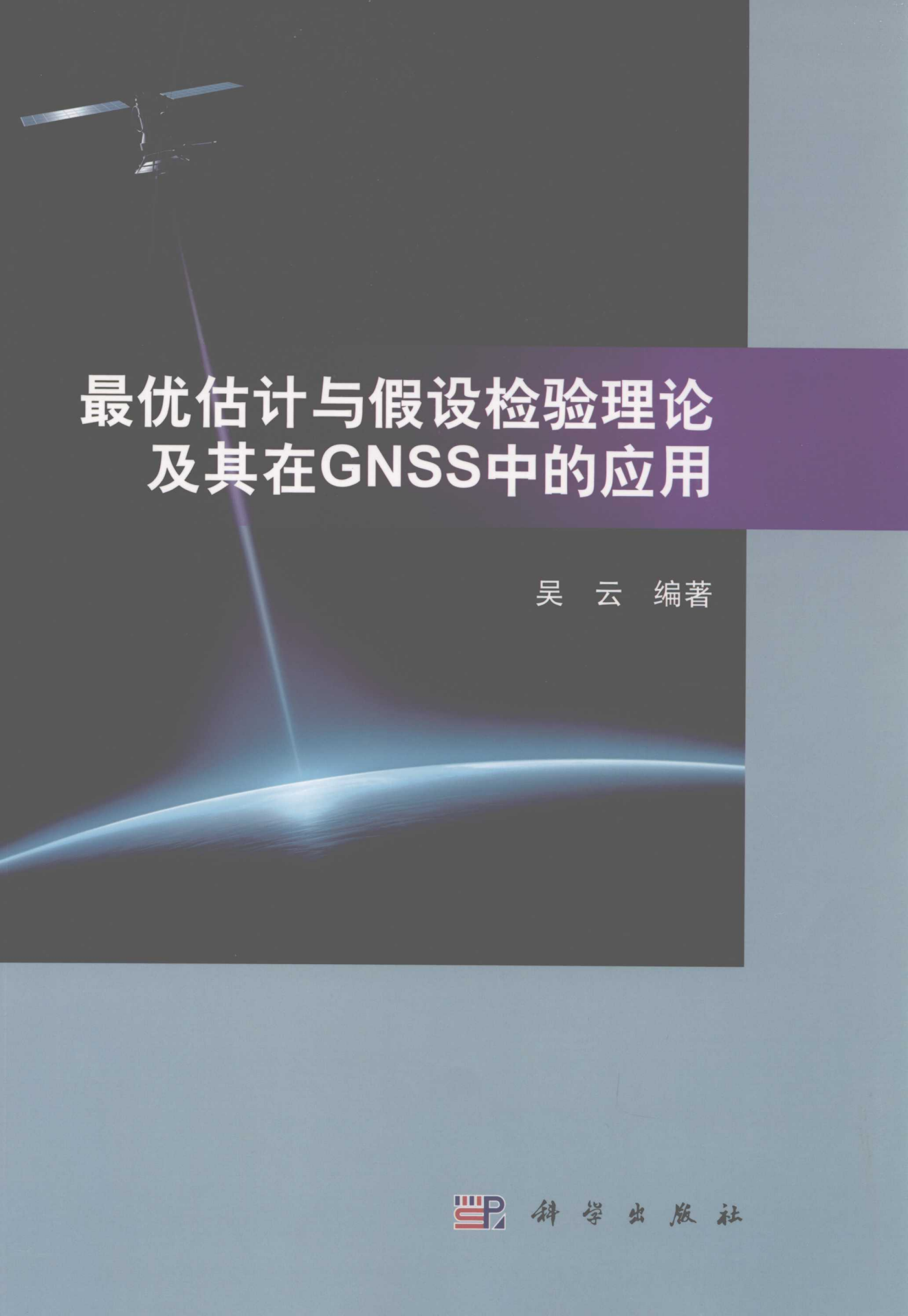




最优估计与假设检验理论 及其在GNSS中的应用

吴 云 编著



科学出版社

最优估计与假设检验理论 及其在 GNSS 中的应用

吴 云 编著

科 学 出 版 社

北 京

版权所有,侵权必究

举报电话:010-64030229;010-64034315;13501151303

内 容 简 介

本书全面系统地阐述了最优估计方法。首先介绍了经典最优估计方法,包括对具体问题的建模、最优估计准则、求解、应用中的问题,各种方法适用的条件和相互之间的关系;然后介绍了动态系统的估计方法,包括动态系统数学模型的建立、线性卡尔曼滤波、预测和平滑方法、线性卡尔曼滤波在实现时的问题和相关对策以及应用算例;最后介绍了假设检验方法和在 GNSS ARAIM 中的应用。

本书介绍的理论基础全面,内容深入浅出,注重理论的应用,可作为导航与测控和测绘工程专业研究生和高年级本科生教材,也可以作为相关领域的工程和科研人员的参考书。

图书在版编目(CIP)数据

最优估计与假设检验理论及其在 GNSS 中的应用 / 吴云编著. —北京:科学出版社, 2015. 11

ISBN 978-7-03-046148-3

I. ①最… II. ①吴… III. ①估计—最佳化理论—应用—卫星导航—全球定位系统②假设检验—应用—卫星导航—全球定位系统 IV. ①P228.4

中国版本图书馆 CIP 数据核字(2015)第 256620 号

责任编辑:张颖兵 杨光华/责任校对:董艳辉 何 念

责任印制:高 嵘 / 封面设计:达 美

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

湖北卓冠印务有限公司印刷

科学出版社发行 各地新华书店经销

*

开本:787×1092 1/16

2015 年 11 月第 一 版 印张:13 1/4

2015 年 11 月第一次印刷 字数:320 000

定价:68.00 元

(如有印装质量问题,我社负责调换)

前 言

用无线电技术对飞机、船舶或其他运动载体进行导航和定位时,无法直接观测到用户的位置,但能够量测到用户与基准线的夹角或者导航台站与用户的距离等观测量,然后用数学的方法来描述观测量与用户位置和姿态,即未知参数的关系,进而求得未知参数的解。但在实际应用中,得到的观测量总是受到测量噪声和误差的干扰,这导致了观测量的不确定性,即观测量是随机变量。此类问题的求解不再是简单的由方程求解未知数的问题,而是如何将有用的信息从被干扰观测值中提取出来,推算参数真值的最佳估值的问题。针对此问题,从统计学发展出来的一系列估计方法称为估计理论。

估计问题的第一步是要建立估计问题的数学模型。数学模型既要考虑到观测值与待估参数的确定性关系,也要考虑到观测值和待估参数的不确定性,即它们的随机特性,然后在数学模型的基础上,根据特定的最优准则推导出参数的估计值。若将系统看做是静态的,在其数学模型中将待估计参数作为静态来描述其与观测值之间的关系,这一类的估计称为静态估计或者参数估计,此类估计也是统计论中的经典估计方法。若将系统看做是动态的,这时参数表示系统连续运动的状态,在其数学模型中不仅建立观测值与参数的关系,而且用状态方程来描述系统的动态过程,这一类估计称为状态估计或者过程估计。

最优估计是在满足一定的假设条件后基于某种准则下的最优估计,不满足假设条件的估计结果会严重偏离实际值而不被发现,所以在估计完成后需要对模型是否满足假设进行检验,估计和假设检验在实际应用中密不可分,都是统计理论中的重要组成部分。

最优估计理论和假设检验的应用非常广泛,如雷达目标状态估计、飞行器轨道的确定、卫星导航和定位、惯性导航、工业过程控制、变形监测、信号分析、计算机图形处理等,可以说最优估计理论和假设检验的应用已经渗透到军事和工业生产的各个领域,更是导航工程专业和测绘工程专业研究生必须掌握的专业理论。

本书从实用角度来介绍最优估计方法和假设检验,旨在帮助从事测绘和导航工程领域的工程人员和高校学生解决工程现实问题。本书无论在介绍经典估计方法还是线性卡尔曼滤波都是从建立数学模型开始,即将现实问题转化为数学模型。在最小二乘和线性卡尔曼滤波的介绍中,每个章节都有算例。通过算例和其中的图表,读者可以认识和体会这些估计方法的性能和在实际应用时所注意的问题。假设检验部分不仅介绍了假设检验的基本理论方法,还详细介绍了多假设检验在 GNSS 接收机的自主完备性监测(GNSS ARAIM)中的应用和具体算法。

本书共分为六章。第 1 章介绍随机变量和随机过程的概念。第 2 章介绍经典最优估计方法。在第二章中首先介绍了参数估计的数学模型,接着阐述了经典最优估计方法:最小二乘估计、极大似然估计、极大验后估计、最小方差估计和线性最小方差估计。其中最小二乘估计是在工程中应用最广泛的估计方法。在此部分,本书以 GNSS 观测值为例,

给出从建模、估计、迭代计算和结果的全过程。通过此算例,读者可以清楚地看到各变量的数值变化,积累对实测数据处理的经验。此外,在此部分中,本书还对几种经典估计方法的异同点和适用条件做了综合的对比分析,有助于读者对这几种经典估计方法的理解和运用。第3章中介绍了动态系统的数学模型的建立。第4章介绍了动态系统的估计方法,即线性卡尔曼滤波、预测和平滑方法并给出算例。平滑的方法有很多,本书给出的是在应用中可以实现并且算法稳定的平滑方法。第5章介绍了卡尔曼滤波在实施中的问题,如计算发散,噪声有色等问题并给出相应的对策和算例。第6章介绍假设检验方法和在 GNSS ARAIM 中的应用。ARAIM 算法用多故障假设来量化 GNSS 的完备性风险,并对多故障进行假设检验以剔除故障卫星来保证 GNSS 用户的安全性,是假设检验理论在 GNSS 中的重要应用。

本书在写作的过程中得到了同行们的大力支持,其中武汉大学测绘学院的姚宜斌教授和李英冰副教授给本书的内容给出了宝贵的意见,特此表示感谢。武汉大学测绘学院的陈立娟博士对本书的部分公式进行了推导和验证,在此对她表示感谢。此外,本书在编写过程中参考了相关领域的大量的专著和论文,在此对这些文献的作者表示感谢。

本书的出版是在国家自然科学基金(41304005)、武汉大学测绘工程专业课程体系综合改革项目(2010018)和“武汉市住房保障和房屋管理局信息化建设公开招标采购项目(智慧房管主数据库设计及一期建库)”的支持下完成的。

由于作者水平有限,书中难免有不足之处,敬请专家和读者批评指正。

编 者

2015 年 4 月

目 录

第 1 章 随机过程	1
1.1 随机变量	1
1.1.1 随机变量的概念	1
1.1.2 随机变量的分布	1
1.1.3 随机变量的数字特征	2
1.1.4 随机变量函数的数学期望和方差	6
1.2 随机过程的概念	8
1.3 随机过程的统计描述	11
1.3.1 随机过程的概率分布	11
1.3.2 随机过程的数字特征	14
1.4 平稳随机过程	16
1.4.1 严格平稳随机过程	17
1.4.2 广义平稳随机过程	18
1.4.3 严格平稳随机过程与广义平稳随机过程的关系	20
1.5 随机过程的各态历经性	20
1.6 典型的随机过程	22
1.6.1 白噪声过程	22
1.6.2 高斯过程	23
1.6.3 高斯白噪声	24
第 2 章 最优估计基础理论	25
2.1 参数估计问题的数学模型	25
2.1.1 数学模型的建立	25
2.1.2 函数模型的线性化	29
2.2 最小二乘估计	33
2.2.1 最小二乘估计	33
2.2.2 最小二乘估计的正交特性	38
2.2.3 附有约束条件的最小二乘估计	39
2.2.3 递推最小二乘估计	41
2.2.4 举例分析	44
2.3 极大似然估计	51
2.3.1 极大似然估计	51
2.3.2 举例分析	51
2.4 极大验后估计	53

2.4.1 极大验后估计	53
2.4.2 举例分析	54
2.5 最小方差估计	56
2.5.1 最小方差估计	56
2.5.2 举例分析	59
2.6 线性最小方差估计	60
2.6.1 线性最小方差估计	60
2.6.2 举例分析	61
2.7 贝叶斯估计	61
2.8 参数估计方法的相互关系	64
第3章 动态系统的数学模型	66
3.1 动态系统的状态方程	66
3.1.1 状态方程	66
3.1.2 状态方程的解	69
3.2 连续线性动态系统的数学模型	74
3.2.1 连续线性动态系统的函数模型	74
3.2.2 连续线性动态系统的随机模型	75
3.3 离散线性系统的数学模型	76
3.3.1 离散线性系统的函数模型	76
3.3.2 离散线性系统的随机模型	79
3.4 动态系统可控性和可测性	80
3.4.1 可控性	81
3.4.2 可测性	83
第4章 线性卡尔曼滤波	87
4.1 卡尔曼滤波概述	87
4.2 线性离散系统卡尔曼滤波	88
4.2.1 基于最小方差准则的推导	89
4.2.2 卡尔曼滤波器的基本递推公式和直观解释	93
4.2.3 卡尔曼滤波的观测值逐次更新法	97
4.2.4 卡尔曼滤波器应用举例	99
4.3 线性离散卡尔曼滤波器的稳定性	105
4.3.1 卡尔曼滤波器的稳定性	105
4.3.2 卡尔曼滤波稳定性的判别条件	106
4.4 线性离散系统的最优预测与平滑	109
4.4.1 线性离散系统的最优预测	109
4.4.2 线性离散系统的最优平滑	111
4.5 线性连续系统的卡尔曼滤波	116

第 5 章 线性卡尔曼滤波的实施	121
5.1 平方根滤波	122
5.1.1 计算误差导致发散的原因	122
5.1.2 正定矩阵的三角分解	123
5.2 UDU^T 分解滤波	132
5.2.1 UDU^T 分解滤波递推公式	133
5.2.2 UDU^T 分解滤波递推的证明	135
5.3 信息滤波	147
5.4 系统噪声与观测噪声相关时的卡尔曼滤波	148
5.5 有色噪声的卡尔曼滤波	150
5.5.1 系统噪声为有色噪声的卡尔曼滤波	150
5.5.2 观测噪声为有色噪声的卡尔曼滤波	151
5.6 自适应的卡尔曼滤波	153
5.6.1 次优无偏极大验后估计器	154
5.6.2 固定窗口的估计方法	158
5.6.3 Sage-Husa 估计方法	159
5.7 扩展的卡尔曼滤波	160
第 6 章 假设检验与 GNSS ARAIM 算法	163
6.1 假设检验的基本概念	163
6.1.1 假设检验	163
6.1.2 假设检验的两类错误	164
6.2 最大功效检验	165
6.2.1 二元假设的最大功效检验	165
6.2.2 多元假设检验	167
6.3 线性模型中的假设检验	169
6.3.1 广义似然比检验	169
6.3.2 线性模型假设检验量的等价形式	171
6.4 GNSS ARAIM 算法	175
6.4.1 GNSS ARAIM 的发展和现状	175
6.4.2 GNSS 导航性能要求	176
6.4.3 ARAIM 算法	178
6.4.5 GNSS ARAIM 的可用性	183
参考文献	184
附录	186
附录 A 矩阵代数基础知识	186
A.1 矩阵的秩	186

A.2	矩阵的迹	186
A.3	特征值和特征向量	186
A.4	广义特征值和特征向量	188
A.5	矩阵的分解	189
A.6	幂等矩阵	191
A.7	矩阵恒等式	191
A.8	矩阵的范数和状态	192
附录 B	向量和矩阵的微分运算	193
B.1	矩阵对变量的微分	193
B.2	函数对向量的微分	194
B.3	矩阵的迹对矩阵的偏导数	194
B.4	特殊函数的微分	195
附录 C	二次型及有关定理	196
C.1	二次型定义	196
C.2	二次型及二次型矩阵的正定性	197
C.3	正定矩阵的判别法	197
C.4	二次型定理	197
附录 D	随机变量的分布及其数字特征	198
D.1	一维随机变量的分布和数字特征	198
D.2	二维随机变量的分布和数值特征	199
D.3	随机变量的边缘分布	201
D.4	随机变量的条件分布	201

第 1 章 随机过程

1.1 随机变量

1.1.1 随机变量的概念

设 E 是随机试验, 其样本空间为 S , 若对每一个样本点 $e \in S$, 都有唯一确定的实数 $X(e)$ 与之对应, 则称 S 上的实值函数 $X(e)$ 是一个随机变量(简记为 X)。随机变量的取值随实验的结果而定, 在实验之前不能预知它取什么值。如果我们对这一对象进行重复观测, 尽管其真值 $\tilde{Z}$ 客观存在, 但由于观测值中有观测误差, 每次观测的观测值并不相同:

$$Z_i = \tilde{Z} + \Delta_i \quad (i = 1, \dots, \ell) \quad (1.1.1)$$

式中: Δ_i 是观测误差。对 $\tilde{Z}$ 的每一次观测其观测误差都不相同, 其数值也无法预料, 所以 Δ_i 是随机变量, 也称为随机误差。随机误差的大小和符号都表现出偶然性, 所以也被称为偶然误差。在 GPS 测量中, 接收机的噪声和对测距码或者相位的有限分辨都导致了 GPS 观测值的误差。从式(1.1.1)可知, 正是由于随机误差的存在, 所以观测值也是随机变量, 如 GPS 接收机在某一时刻观测得到的 GPS 卫星与接收机之间的距离、对某物体的方位角观测和卫星在空间运行时某一时刻的位置等都是随机变量。

随机误差不是干扰观测值的唯一的误差, 有时观测值还被系统误差和粗差干扰。系统误差大小和数值正负上会表现出系统性, 或者在观测过程中按一定的规律变化, 如 GPS 观测值中电离层和对流层对信号的延迟造成的观测误差。电离层和对流层延迟误差可以通过经验模型对其进行一定程度上的改正, 从而减小系统误差对观测值的影响。粗差指在数值上表现为较大的误差, 如在 GPS 导航中由于卫星钟的跳变引起观测误差等。从随机特性上看, 粗差出现的概率非常小, 所以通常用概率统计的方法发现并剔除受粗差干扰的异常观测值。

在本书中, 如果没有特别说明, 观测误差就是指的随机误差。

1.1.2 随机变量的分布

有些随机变量, 它全部取值是有限个或者可列无限多个, 这种随机变量叫离散型随机变量, 如骰子出现的点数等。离散型随机变量的取值对应一个确定的概率, 这个概率可以由它的概率分布获得。离散型随机变量常见的概率分布有二项分布(binomial distribution)、伯努利分布(Bernoulli distribution)和泊松分布(Poisson distribution)等。

有的随机变量可能的取值充满一个区域, 但是无法按照一定次序一一列举出来, 这样的随机变量是连续性随机变量。对于连续型随机变量就不能像离散型随机变量那样可以用分布规律来描述它, 连续型随机变量取任意指定的实数值的概率为零, 其概率分布由一

个连续的函数来描述,即概率分布函数

$$F(x) = \int_{-\infty}^x f(u) du \quad (1.1.2)$$

其中: $f(u)$ 为随机变量的概率密度函数。连续型随机变量 X 的概率分布常用的有正态分布(normal/gaussian distribution)、指数分布(exponential distribution)和均匀分布(uniform distribution)等。随机变量 X 的正态分布概率密度函数为

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left\{-\frac{(x - \mu_x)^2}{2\sigma_x^2}\right\} \quad (1.1.3)$$

其中: μ_x 为随机变量 X 的数学期望; σ_x^2 为随机变量 X 的方差。数学期望和方差是随机变量 X 的重要数值特征,决定了随机变量 X 的分布情况,记为 $X \sim N(\mu_x, \sigma_x^2)$ 。

就随机误差个体而言,其数值是不能预知的,但从大样本分析来看,具有统计规律性,我们可以通过大量的试验得到观测误差的经验分布。如我们对三角形的内角和进行多次观测,可得到观测误差(误差真值)

$$\Delta_i = w_i - 180^\circ \quad (i = 1, \dots, \ell) \quad (1.1.4)$$

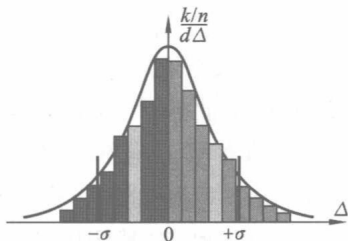


图 1.1 随机误差分布直方图和分布曲线

式中: w_i 是三角形内角和的观测值。在得到大样本的观测误差后,可以用直方图描述观测误差的分布情况(图 1.1)。图 1.1 中的横坐标为随机误差的数值,纵坐标为在区间 $(\Delta, \Delta + d\Delta)$ 出现的随机误差的频率与区间长度 $d\Delta$ 的比值。若观测次数无限增大,并将误差区间长度 $d\Delta$ 无限缩小,那么图中的直方图就变成了一条平滑的曲线。这条曲线就是观测误差的经验分布。根据概率论的中心极限定理,随着

观测次数的增大,其分布无限接近正态分布,所以正态分布是观测误差的理论分布。在对实际观测数据进行处理时,我们通常假设观测随机误差服从期望为零的正态分布 $N(0, \sigma^2)$ 。

除了观测随机误差外,在自然现象和社会现象中,大量的随机变量都服从或者近似服从正态分布,如海洋波浪的高度和电子管的热噪声电流等都服从正态分布。

1.1.3 随机变量的数字特征

分布函数能够完整地描述随机变量的统计特性,但在一些实际问题中,无法得到随机变量的分布,或者不需要去全面考察随机变量的变化情况,而只需要知道随机变量的某些特征,如对某一段距离进行重复量测,人们关心的是其量测值的平均值和量测值与平均值的偏离程度,这些数值虽然不能完整地描述随机变量,但能够描述随机变量在某些方面的重要特征。因此,本节将介绍随机变量常用的数字特征:数学期望、方差和相关系数,以及如何由随机变量的数字特征求得随机变量函数的数字特征。

1. 随机变量的数学期望

数学期望又称为均值。对于连续随机变量,其数学期望的定义为

$$E(X) = \mu_x = \int_{-\infty}^{+\infty} x f(x) dx \quad (1.1.5)$$

如果随机变量 X 服从正态分布,根据式(1.1.5)和式(1.1.3)的分布可求得其数学期望为 μ_X ,它是随机变量 X 分布所在的中心位置。对于离散随机变量,其数学期望的定义为

$$\mu_X = \sum_{i=1}^{\infty} x_i P(x_i) \quad (1.1.6)$$

其中: $P(x_i)$ 是随机变量 X 取值为 x_i 的概率。从式(1.1.6)看出,离散随机变量的数学期望为离散数值 $x_i (i=1, \dots, n)$ 的“加权”平均值(weighted mean)。如果 X 在其值域内每个数值取值概率相等,即为 $1/n$,式(1.1.6)即为随机变量 X 的算数平均值。

如果视 $x_i (i=1, \dots, n)$ 为连续随机变量 X 的有限样本,即可由式(1.1.6)求得随机变量 X 的样本平均值,它是对总体分布期望 μ_X 的无偏估计。

2. 随机变量的方差

在求得平均值后,还需要进一步求得 x_i 与平均值的偏离程度,偏离程度小,说明 x_i 数值稳定。随机变量 X 与它的均值 $E(X)$ 的偏离程度定义为

$$\text{var}(X) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f(x) dx = E[(X - E(X))^2] \quad (1.1.7)$$

即随机变量 X 的二阶中心矩, $\text{var}(X)$ 称为方差, $\sqrt{\text{var}(X)}$ 称为标准差。方差表达了随机变量 X 的取值或者样本与均值的离散程度,它是衡量随机变量 X 或者样本之间离散程度的一个尺度。

如果随机变量 X 服从正态分布,根据式(1.1.3)的分布和式(1.1.7),可求得其方差为 σ_X^2 ,它是随机变量 X 概率密度函数的拐点,决定了 X 分布的形状。图 1.2 所示随机变量 X_1 和 X_2 的分布情况。显然与 X_2 比较, X_1 的取值与其期望的偏离程度小,更集中地分布在期望附近,其方差为 $\sigma_{X_1}^2 < \sigma_{X_2}^2$ 。

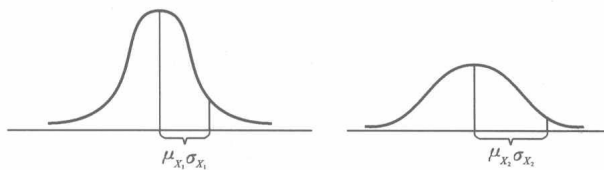


图 1.2 随机变量 X_1 和 X_2 的分布情况

随机变量与其期望的偏离程度可以用出现在期望附近的区间的概率来表示,服从正态分布的随机变量 X 在区间 $[\mu_x - m\sigma_x, \mu_x + m\sigma_x]$ 概率为

$$\begin{aligned} P(\mu - m\sigma_x < X < \mu + m\sigma_x) &= \int_{\mu - m\sigma_x}^{\mu + m\sigma_x} f(x) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mu - m\sigma_x}^{\mu + m\sigma_x} \exp\left\{-\frac{1}{2\sigma^2} (x - \mu)^2\right\} dx \end{aligned} \quad (1.1.8)$$

取 $m = 1, 2, 3$ 时,其概率分布为

$$\begin{aligned} P(\mu - \sigma_x < X < \mu + \sigma_x) &\approx 68.3\% \\ P(\mu - 2\sigma_x < X < \mu + 2\sigma_x) &\approx 95.5\% \\ P(\mu - 3\sigma_x < X < \mu + 3\sigma_x) &\approx 99.7\% \end{aligned} \quad (1.1.9)$$

上式给出了随机变量 X 在 $(\mu - m\sigma_x, \mu + m\sigma_x)$ 可信程度,即一定置信度下的置信区间,如

其中的第二式表明随机变量 X 在均值附近 2 倍标准差范围内出现的概率为 95.5%, 超出此范围的概率仅为 4.5%, 如果某个观测值在这个区间之外, 可将此观测值视为异常观测值(粗差观测值)。因此在工程中, 常将 $2\sigma_X$ 或者 $3\sigma_X$ 作为剔除异常观测值的门限值。

在现实中随机变量 X 的方差可由样本 $x_i (i = 1, \dots, n)$ 计算得到

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \quad (1.1.10)$$

其中: $\bar{X}$ 是样本平均值; S_X 称为样本标准差, 它是对总体分布标准差 σ_X 的无偏估计。当样本个数足够大时, 式(1.1.10)中分母部分可由 n 代替, 计算结果并无明显差异。

在工程应用中, 常用 S_X 来评价样本(观察值或者观测值)质量的好坏, 即精度(precision)的好坏。如图 1.2 所示, 随机变量 X_1 的精度好于随机变量 X_2 。

在某些情况下, 研究对象的真值(参考值, reference value), 即 $\tilde{X}$ 已知的情况下, 可用均方差 MSE (mean square error) 衡量样本与真值的离散程度, 即

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{X})^2 \quad (1.1.11)$$

均方差 MSE 评价的是样本与真值(参考值)的符合程度, 它反映的是观察值的准确度(accuracy), 能更加全面地评价观测值质量的好坏。 $\sqrt{MSE}$ 即为 RMS (root of mean square)。可以证明均方差与方差的关系为

$$MSE(X) = \sigma_X^2 + (E(X) - \tilde{X})^2 \quad (1.1.12)$$

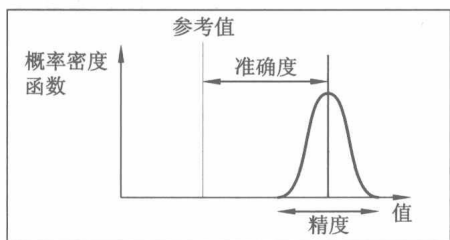


图 1.3 精度与准确度的关系

即 X 的均方误差等于 X 的方差(偶然方差)加上平均值与真值偏差的平方。图 1.3 反映了精度和准确度在样本分布上的差异。方差 σ_X^2 反映的是观测值与其自身平均值(期望)之间的离散程度, 平均值一般在其出现的概率最大的位置; 而 MSE 不仅反映了观测值与其期望之间的离散程度, 而且也反映了观测值与已知参考值之间的差异。观察式(1.1.10), 当观测值受到系统误差的干扰时, 其平均值也受到

系统误差的影响, 所以在计算 S_X 时, 系统误差被抵消, S_X 并不能全面反映出观测值是否受到系统误差的干扰。所以与 σ_X^2 比较, MSE 更能客观的反映出观测值的质量。

3. 随机变量的协方差和相关系数

对于二维随机变量 (X_1, X_2) , 我们除了讨论他们各自的数学期望和方差外, 还需要讨论它们之间关系的数字特征, 即协方差和相关系数。二维随机变量 (X_1, X_2) 的协方差定义为

$$\text{cov}(X_1, X_2) = \sigma_{X_1 X_2} = E[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})] \quad (1.1.13)$$

由随机变量 (X_1, X_2) 的样本可以计算样本协方差

$$S_{X_1 X_2} = \sum_{i=1}^n (x_{1,i} - \bar{X}_1)(x_{2,i} - \bar{X}_2) P(x_{1,i}, x_{2,i}) \quad (1.1.14)$$

其中 $P(x_{1,i}, x_{2,i})$ 为

$$P(x_{1,i}, x_{2,i}) = P(X_1 = x_{1,i}, X_2 = x_{2,i}) \quad (1.1.15)$$

若随机变量 X_1 和 X_2 都服从正态分布, 并记 $\mathbf{X} = (X_1, X_2)$, 随机向量 $\mathbf{X}$ 的概率密度函数为

$$f(\mathbf{X}) = \frac{1}{(2\pi)^{\frac{n}{2}} |D_{\mathbf{X}}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} (\mathbf{X} - \mu_{\mathbf{X}})^T D_{\mathbf{X}}^{-1} (\mathbf{X} - \mu_{\mathbf{X}})\right\} \quad (1.1.16)$$

上式即为二维随机变量 (X_1, X_2) 的联合概率密度函数 $f(x_1, x_2)$ 。其中

$$\mu_{\mathbf{X}} = \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} \quad (1.1.17)$$

$$D_{\mathbf{X}} = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 \end{bmatrix} \quad (1.1.18)$$

$D_{\mathbf{X}}$ 称为二维随机变量 (X_1, X_2) 的协方差矩阵, 其中 $\sigma_{X_1}^2$ 为随机变量 X_1 的方差, $\sigma_{X_1 X_2}$ 为随机变量 X_1 和 X_2 的协方差。图 1.4 即为正态随机变量 (X_1, X_2) 的联合概率密度函数 $f(\mathbf{X})$, (X_1, X_2) 出现在区域 Ω 的概率为以区域 Ω 为底(图 1.5), $f(\mathbf{X})$ 为顶的柱体的体积, 即 $f(x)$ 在区域 Ω 的积分。

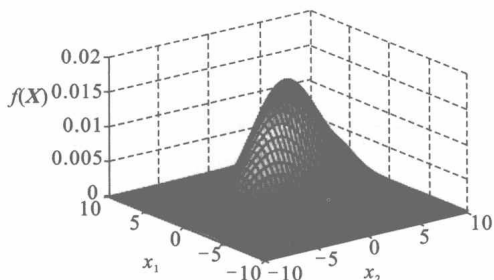


图 1.4 正态随机变量 (X_1, X_2) 的联合概率密度函数 $f(x_1, x_2)$

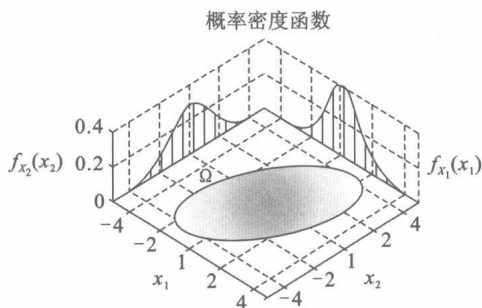


图 1.5 正态随机变量 (X_1, X_2) 的边缘分布和区域 Ω

方差和协方差具有功率的量纲, 为了消除量纲, 做如下处理

$$\rho_{X_1 X_2} = \frac{\sigma_{X_1 X_2}}{\sigma_{X_1} \sigma_{X_2}} \quad (1.1.19)$$

$\rho_{X_1 X_2}$ 即为随机变量 X_1 与 X_2 的相关系数。由样本计算的相关系数为

$$\rho_{X_1 X_2} = \frac{S_{X_1 X_2}}{S_{X_1} S_{X_2}} \quad (1.1.20)$$

如图 1.6 为随机变量 X_1 和 X_2 的样本 $x_{1,i}$ 和 $x_{2,i}$, 图 1.6 表明 X_1 和 X_2 有较强的负相关性, 当 X_1 的样本值增大时, X_2 的样本值减小, 由其样本计算得到的相关系数为 -0.6 。

相关系数的取值范围为

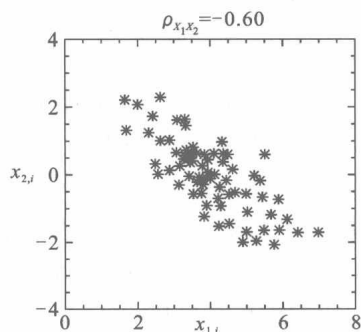


图 1.6 随机变量 X_1 和 X_2 的样本 $x_{1,i}$ 和 $x_{2,i}$

$$|\rho_{X_1 X_2}| \leq 1 \quad (1.1.21)$$

若 $\rho_{X_1 X_2} = 0$, 则表明随机变量 X_1 与 X_2 不相关(uncorrelated), 即 $\text{cov}(X_1, X_2) = 0$ 。这里需要注意的是, 随机变量间的相关性和独立性是两个不同的概念。两个随机变量相互独立(statistical independence) 是指随机变量 X_1 与 X_2 的分布如果满足

$$f(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \quad (1.1.22)$$

上式中的 $f_{X_1}(x_1)$ 和 $f_{X_2}(x_2)$ 分别为 x_1 和 x_2 的边缘分布。如果上式成立, 那么 X_1 与 X_2 相互独立, 没有任何关系, 自然也不会相关, 所以有

随机变量 X_1 与 X_2 相互独立 $\Rightarrow$ 随机变量 X_1 与 X_2 不相关

但是, 不相关并不意味着随机变量 X_1 与 X_2 相互独立:

随机变量 X_1 与 X_2 不相关 $\nRightarrow$ 随机变量 X_1 与 X_2 相互独立

在特殊情况下, 如当 (X_1, X_2) 服从二维正态分布, 由随机变量 X_1 与 X_2 不相关可以得到 $\text{cov}(X_1, X_2) = 0$, 即可以求得 $f(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$, 随机变量 X_1 与 X_2 相互独立。这意味着当 (X_1, X_2) 服从正态分布时, X_1 与 X_2 不相关和 X_1 与 X_2 相互独立是等价的。

对于多维随机变量

$$\mathbf{X} = [X_1 \quad X_2 \quad \cdots \quad X_n]^T \quad (1.1.23)$$

其期望值为

$$\boldsymbol{\mu}_X = \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \vdots \\ \mu_{X_n} \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{bmatrix} = E(\mathbf{X}) \quad (1.1.24)$$

协方差矩阵为

$$\mathbf{D}_{XX} = E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^T] = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} & \cdots & \sigma_{x_1 x_n} \\ \sigma_{x_2}^2 & \cdots & \sigma_{x_2 x_n} \\ \vdots & & & \\ \text{symmetric} & & & \sigma_{x_n}^2 \end{bmatrix} \quad (1.1.25)$$

1.1.4 随机变量函数的数学期望和方差

在实际应用中, 需要研究的对象 X 往往无法直接测量, 但可以建立其与观测量的函数关系, 因此可以根据可量测量 Z 来推算研究对象。现假设 Z 与随机变量 X 有函数关系为

$$Z = f(X_1, X_2, \dots, X_n) \quad (1.1.26)$$

这样 Z 也是随机变量。若已知随机变量 X 的随机特性, 我们希望通过上面的函数关系得到随机变量 Z 的随机特性。本部分将讨论如何根据随机变量 X 的期望和方差得到随机变量的 Z 的期望和方差。

1. 随机变量线性函数的数学期望和方差

为不失一般性

$$\mathbf{X} = [X_1 \quad X_2 \quad \cdots \quad X_n]_{n,1}^T \quad (1.1.27)$$

设有线性函数

$$Z = k_1 X_1 + k_2 X_2 + \cdots + k_n X_n + k_0 \quad (1.1.28)$$

可以将上式表示为

$$Z = \mathbf{KX} + k_0 \quad (1.1.29)$$

其中

$$\mathbf{K} = (k_1 \quad k_2 \quad \cdots \quad k_n) \quad (1.1.30)$$

由期望的定义可得到随机变量 Z 的期望

$$E(Z) = E(\mathbf{KX} + k_0) = \mathbf{KE}(\mathbf{X}) + k_0 = \mathbf{K}\boldsymbol{\mu}_X + k_0 \quad (1.1.31)$$

根据方差的定义可以得到

$$D_{ZZ} = \sigma_Z^2 = E[(Z - E(Z))(Z - E(Z))^T] \quad (1.1.32)$$

将式(1.1.31) 带入上式得到

$$\begin{aligned} D_{ZZ} &= \sigma_Z^2 = E[(\mathbf{KX} - \mathbf{K}\boldsymbol{\mu}_X)(\mathbf{KX} - \mathbf{K}\boldsymbol{\mu}_X)^T] \\ &= \mathbf{KE}[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^T]\mathbf{K}^T \\ &= E[\mathbf{K}(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^T\mathbf{K}^T] \end{aligned} \quad (1.1.33)$$

所以,随机变量 Z 的方差为

$$D_{ZZ} = \sigma_Z^2 = \mathbf{KD}_{XX}\mathbf{K}^T \quad (1.1.34)$$

已知随机变量的方差,求得随机变量函数的方差也称为误差传播定律。

2. 随机变量线性函数的协方差

若有另一个研究对象 Y 也是同一组随机变量 $\mathbf{X}$ 的函数

$$Y = g_1 X_1 + g_2 X_2 + \cdots + g_n X_n + g_0 \quad (1.1.35)$$

令 $\mathbf{G} = (g_1 \quad g_2 \quad \cdots \quad g_n)$, 式(1.1.35) 可以表示为

$$Y = \mathbf{GX} + g_0 \quad (1.1.36)$$

由于随机变量 Y 与随机变量 Z 是同一组随机变量 $\mathbf{X}$ 的函数, Y 与 Z 之间必然存在着相关关系。由协方差的定义可求得

$$\text{cov}(Z, Y) = \sigma_{ZY} = E[(Z - \mu_Z)(Y - \mu_Y)^T] \quad (1.1.37)$$

将式(1.1.29) 和式(1.1.36) 带入上式,可以得到

$$\sigma_{ZY} = \mathbf{KD}_{XX}\mathbf{G}^T \quad (1.1.38)$$

得到随机变量 Y 与随机变量 Z 的各自的方差和协方差和后,可以求得 Y 与 Z 的相关系数。

3. 随机变量非线性函数的数学期望和方差

在实际应用中,很多研究对象是观测值的非线性函数,这时可以将非线性函数通过泰勒级数展开,然后舍弃高阶项得到函数 Z 与 $\mathbf{X}$ 的线性函数形式,从而将复杂的数学计算简单化。首先给随机变量 $\mathbf{X}$ 赋以初始值

$$\mathbf{X}^0 = [X_1^0 \quad X_2^0 \quad \cdots \quad X_n^0]^T \quad (1.1.39)$$

将函数 Z 在 $\mathbf{X}^0$ 处展开为泰勒级数

$$Z = f(X_1^0 \quad X_2^0 \quad \cdots \quad X_n^0) + \left(\frac{\partial h}{\partial X_1}\right)_0 (X_1 - X_1^0) + \left(\frac{\partial h}{\partial X_2}\right)_0 (X_2 - X_2^0) + \cdots$$

$$+ \left(\frac{\partial h}{\partial X_n} \right)_0 (X_n - X_n^0) + (\text{二阶项和高阶项}) \quad (1.1.40)$$

$\mathbf{X}$ 与 $\mathbf{X}^0$ 的数值差异越小, 其二阶项和高阶项就越小, 因此可以将二阶项和高阶项忽略(忽略项对函数的影响将在后面章节叙述), 这样函数 Z 可表达为

$$Z = \left(\frac{\partial f}{\partial X_1} \right)_0 X_1 + \left(\frac{\partial f}{\partial X_2} \right)_0 X_2 + \cdots + \left(\frac{\partial f}{\partial X_n} \right)_0 X_n + f(X_1^0 \quad X_2^0 \quad \cdots \quad X_n^0) - \sum_{i=1}^n \left(\frac{\partial f}{\partial X_i} \right)_0 X_i^0 \quad (1.1.41)$$

令

$$\mathbf{K} = [k_1 \quad k_2 \quad \cdots \quad k_n] = \left[\left(\frac{\partial h}{\partial X_1} \right)_0 \quad \left(\frac{\partial h}{\partial X_2} \right)_0 \quad \cdots \quad \left(\frac{\partial h}{\partial X_n} \right)_0 \right] \quad (1.1.42)$$

和

$$k_0 = f(X_1^0 \quad X_2^0 \quad \cdots \quad X_n^0) - \sum_{i=1}^n \left(\frac{\partial f}{\partial X_i} \right)_0 X_i^0 \quad (1.1.43)$$

函数 Z 即为

$$Z = k_1 X_1 + k_2 X_2 + \cdots + k_n X_n + k_0 \quad (1.1.44)$$

这样非线性函数 Z 就转化为了线性函数, 这个过程称为非线性函数的线性化。式(1.1.44)与式(1.1.28)形式完全一致, 因此可由式(1.1.31)和式(1.1.34)求得随机函数 Z 的期望和方差。

从以上的推导可以看出, 在求非线性函数 Z 的方差时, 只需要求得函数 Z 在 $\mathbf{X}^0$ 处的一阶偏导数 $k_i (i = 1, \cdots, n)$, 所以可以跳过泰勒级数的展开步骤, 直接来求函数 Z 的一阶偏导数

$$dZ = \left(\frac{\partial f}{\partial X} \right)_0 dX_1 + \left(\frac{\partial f}{\partial X_2} \right)_0 dX_2 + \cdots + \left(\frac{\partial f}{\partial X_n} \right)_0 dX_n = \mathbf{K} d\mathbf{X} \quad (1.1.45)$$

组成矩阵 $\mathbf{K}$, 即一阶的雅克比矩阵, 然后利用式(1.1.34)得到随机函数 Z 的方差。

1.2 随机过程的概念

从前面的学习可以知道, 随机实验所有可能出现的结果可以用随机变量的取值来表示。但是, 如果研究对象是随时间变化的, 每次试验(观测)所得到的观测过程都不相同, 他们是时间 t 的不同函数, 试验(观测)前又不能预知这次试验(观测)会出现什么结果, 没有确定的变化规律, 这样的过程称为随机过程。对连续时间的随机过程进行抽样得到的序列称为离散时间随机过程, 或简称为随机序列。连续时间的随机过程用 $X(t)$ 表示, 随机序列用 $X(n)$ 表示。

下面来看几个随机过程的例子。

例 1.1 正弦型随机相位信号

$$X(n, \Phi) = A \cos(\omega_0 n + \Phi)$$

其中: A 和 ω_0 为常数; Φ 为 $(-\pi, \pi)$ 上均匀分布的随机变量。由于起始相位 Φ 是一个连续型的随机变量, 它的取值范围为 $(-\pi, \pi)$ 。在观测信号 $X(n, \Phi)$ 之前, 并不能预知 Φ 究竟取何值, 因此, 我们也不能预知 $x(n)$ 究竟取哪个样本。对于固定的 $\varphi_i (-\pi < \varphi_i < \pi)$, 对